

Europäisches Patentamt  
European Patent Office  
Office européen des brevets



(11) **EP 1 083 543 A2**

(12) **EUROPÄISCHE PATENTANMELDUNG**

(43) Veröffentlichungstag:  
14.03.2001 Patentblatt 2001/11

(51) Int. Cl.<sup>7</sup>: **G10L 15/20**

(21) Anmeldenummer: 00117264.2

(22) Anmeldetag: 16.08.2000

(84) Benannte Vertragsstaaten:  
**AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU  
MC NL PT SE**  
Benannte Erstreckungsstaaten:  
**AL LT LV MK RO SI**

(72) Erfinder:  
• Tontch, Kilian, Dipl.-Ing.  
38440 Wolfsburg (DE)  
• Schaaf, Klaus, Dr.  
38116 Braunschweig (DE)  
• Wengelnik, Heino, Dr.  
38440 Wolfsburg (DE)

(30) Priorität: 08.09.1999 DE 19942868

(71) Anmelder:  
Volkswagen Aktiengesellschaft  
38436 Wolfsburg (DE)

(54) **Verfahren zum Betrieb einer Mehrfachmikrofonanordnung in einem Kraftfahrzeug zur sprachgesteuerter Befehleingabe**

(57) Die Erfindung betrifft ein Verfahren zum Betrieb einer Mehrfachmikrofonanordnung in einem Kraftfahrzeug, bei welchem die Einzelsignale der Einzelmikrofone erfaßt und elektronisch nachbearbeitet werden, sowie eine Mehrfachmikrofonanordnung selbst, gemäß Oberbegriff der Patentansprüche 1 und 7. Um hierbei zu erreichen, daß Sprachbefehle bei sprachgenerierten Eingaben besser und sicherer erkannt werden können, ist erfindungsgemäß vorgeschlagen, daß zur präzisen Ansteuerung einer sprachgesteuerten Befehleingabe von Steuer- oder Betätigungsbefehlen, oder zum präzisen Betrieb der Telefonanlage die Signale der Einzelmikrofone simultan und im Zeitbezug zueinander erfaßt und analysiert werden, und daß aus den daraus gewonnenen Parametern und dem Kohärenzbezug der Einzelsignale die Sprachbefehle vor Ansteuerung der Befehleingabe auf Plausibilität geprüft werden.

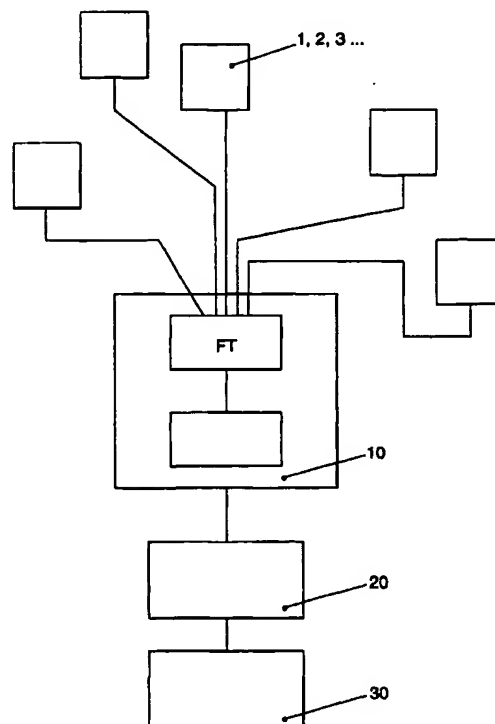


FIG. 1

EP 1 083 543 A2

## Beschreibung

[0001] Die Erfindung betrifft ein Verfahren zum Betrieb einer Mehrfachmikrofonanordnung in einem Kraftfahrzeug, bei welchem die Einzelsignale der Einzelmikrofone erfaßt und elektronisch nachbearbeitet werden, sowie eine Mehrfachmikrofonanordnung selbst, gemäß Oberbegriff der Patentansprüche 1 und 7.

[0002] Die Platzierung von Mikrofonen im Kraftfahrzeug kann mehrere Aufgaben haben. Zum einen sind für den Betrieb von Kommunikationssystemen, wie Bordtelefon und dgl. mehr, Mikrofone notwendig. Überdies gibt es jedoch auch sprachunterstützte Eingabesysteme, mit denen Schalt- oder Betätigungsbefehle sprachgesteuert, also ohne Handhabung von Betätigungselementen ausgelöst werden können. In größeren Fahrzeugen, wie bspw. in Bussen, dienen Mikrofonanordnungen zur Unterstützung von Kommunikation innerhalb des Fahrzeuges.

[0003] Weiterhin sind neuere Anwendungen möglich, bei denen über Mikrofone Schall innerhalb des Fahrzeuges erfaßt wird und elektronisch in der Form verarbeitet wird, daß daraus Gegenbeschallungssignale zur destruktiven Interferenz von Geräuschen innerhalb des Fahrzeuges eingesetzt werden.

[0004] So ist der JP 7 176 178 eine Mehrfachmikrofonanordnung in Kraftfahrzeugen bekannt, welche zur Erfassung der Stimmen aller Mitfahrer dient. Dabei wird erreicht, daß beispielsweise bei in Betrieb genommener Audioanlage eine Leise- oder Stummschaltung der Musik erfolgt, immer dann, wenn einer der Fahrzeuginsassen spricht.

[0005] Aus der EP 0 721 178 A2 ist ein Multikanalkommunikationssystem bekannt, welches zur Unterdrückung von Echos und Störgeräuschen bei der Kommunikation bzw. während der Kommunikation dient.

[0006] Aus der US 3 784 747 ist eine Anordnung bekannt, bei welcher die Unterdrückung von Störsignalen bei der Verwendung von Mehrfachmikrofonanordnungen als solches im Vordergrund steht. Hierbei werden Rückkopplungseffekte und dgl. behandelt. Ein weiteres Geräuschunterdrückungsverfahren ist aus der EP 0 729 288 A2 bekannt.

[0007] Allen diesen Verfahren und Anordnungen ist jedoch eine Problematik grundlegend. Die Platzierung von Mikrofonen fixiert in allen diesen Fällen den Ort der Schallaufnahme innerhalb des Kraftfahrzeuges. Befindet sich nun eine Sprachquelle, d. h. ein Fahrzeuginsasse in entsprechender Entfernung zum besagten Mikrofon, so nimmt dasselbe um noch einigermaßen sensitiv Sprache aufnehmen zu können, natürlich auch Fahrgeräusche oder Geräusche innerhalb des Fahrzeuges auf. Außerdem erhöhen sich Rückkopplungseffekte naturgemäß bei einer solchen Situation.

[0008] Aus der EP 0 660 639 A1 ist daher eine mechanische Ausrichtbarkeit eines Einzelmikrofones in

einem Kraftfahrzeug angegeben. Das Mikrofon ist dabei in einer Art Kugelgelenk angeordnet, welches in verschiedene Richtungen drehbar sein soll, um eine gewisse Richtcharakteristik gewährleisten zu können. Jedoch auch hierbei ergibt sich nachteiligerweise, daß ein so installiertes Einzelmikrofon zwar in geringfügigem Maße gerichtet werden kann, aber die Schallwege zu unterschiedlichen Schallquellen durchaus unterschiedlich groß sein können. Es ist damit auch keine gleichbleibende Qualität der Sprachaufnahme möglich.

[0009] Vor dem Hintergrund sprachgenerierter Eingaben bei Kraftfahrzeugen kommt jedoch der Qualität der ankommenden Sprachsignale eine wesentliche Bedeutung zu. Eine schlechte Spracherkennung führt somit zu einer fehlerbehafteten Betätigung der sprachgesteuerten Funktionen, was im übrigen auch die Fahr-sicherheit beeinträchtigen kann. Dies insbesondere deshalb, weil Fehlfunktionen den Fahrzeugführer immer deutlich vom Verkehrsgeschehen ablenken.

[0010] Der Erfindung liegt somit die Aufgabe zugrunde, sowohl ein Verfahren zum Betrieb einer Mehrfachmikrofonanordnung, sowie auch eine Mehrfachmikrofonanordnung selbst dahingehend zu verbessern, daß Sprachbefehle immer optimal aufgenommen werden können.

[0011] Die gestellte Aufgabe ist bei einem Verfahren der gattungsgemäßen Art erfindungsgemäß durch die kennzeichnenden Merkmale des Patentanspruches 1 gelöst.

[0012] Weitere vorteilhafte Ausgestaltungen des erfindungsgemäßen Verfahrens sind in den nachfolgenden abhängigen Ansprüchen 2-4 angegeben.

[0013] Hinsichtlich einer Mehrfachmikrofonanordnung selbst ist die gestellte Aufgabe erfindungsgemäß durch die kennzeichnenden Merkmale des Patentanspruches 5 gelöst. Weitere vorteilhafte Ausgestaltungen hierzu sind in den übrigen Ansprüchen angegeben.

[0014] Das erfindungsgemäße Verfahren besteht hierbei darin, daß zur präziseren Ansteuerung einer sprachgesteuerten Eingabe von Steuer- oder Betätigungsbefehlen, die Signale der Einzelmikrofone simultan und im Zeitbezug zueinander erfaßt und bspw. fouriertransformiert werden, und daß aus den daraus gewonnenen bzw. gewinnbaren Parametern und dem Kohärenzbezug der Einzelsignale selbst, die Sprachsignale vor Ansteuerung der Befehlseingabe auf Plausibilität geprüft werden.

[0015] Wesentlich ist hierbei, daß mehrere Mikrofone verwendet werden. Diese sind an verschiedenen Orten im Fahrzeug weitestgehend asymmetrisch platziert. Durch diese asymmetrische Verteilung ist es unwahrscheinlich, daß die Sprachquelle dann zu irgendeinem Zeitpunkt gleich weit zu allen Einzelmikrofonen entfernt ist. Wichtig ist hierbei jedoch, daß die Signale aller Einzelmikrofone ausgewertet werden; also auch solche, die näher oder weiter entfernt von der Sprachquelle platziert sind. Dadurch können die Einzelmikrofonsignale simultan und in einem Zeitbezug zuein-

ander erfaßt werden. Durch unterschiedliche Entfernungen zum Mikrofon ergeben sich natürlich auch unterschiedliche Laufzeiten. Schon durch die besagte Laufzeiterfassung der einzelnen Mikrofonsignale ist es möglich, die Sprachquelle als solches räumlich zu orten. Die erfaßten Einzelsignale der Mikrofone werden dabei auch jeweils separat voneinander analysiert d.h. bspw. fouriertransformiert.

**[0016]** Im Falle einer Signalanalyse mit Fouriertransformation findet diese dabei vom Frequenzraum in den Phasenraum statt. Aus den Signalen, und den daraus gewonnenen Parametern sowie dem Kohärenzbezug der Einzelsignale zueinander, können somit die jeweiligen Sprachquellen, die ja unterschiedlich sein können, jeweils geortet werden. Die Fülle der Parameter als solche eröffnet die Möglichkeit vor Ansteuerung der Befehlseingabe eine Plausibilitätsprüfung des ankommenden Sprachbefehles vorzunehmen.

**[0017]** In weiterer vorteilhafter Ausgestaltung der Erfindung ist angegeben, daß zusätzlich aus den Signalparametern und dem Kohärenzbezug eine Richtcharakteristik ermittelt wird, mit Hilfe welcher der virtuelle Mikrofonort der tatsächlichen Sprachquelle nachgeführt wird. Durch eine solche Ortung ist die Mikrofonsignalauswertung als solches optimierbar, wodurch die Sprachbefehle als solches auch besser qualifiziert werden können.

**[0018]** In vorteilhafter Ausgestaltung der Erfindung ist es möglich, daß aus Abprüfung der Plausibilität und virtueller Lokalisierung der Sprachquelle übrige, delokale Geräuschquellen durch Subtraktion eliminiert werden können. Dabei kommt die Ermittlung des virtuellen Mikrofonortes zur Wirkung. Die ankommende Sprachqualität ist dahingehend verbesserbar, daß auch die Nebengeräuschquellen qualifiziert erkannt werden können. Dies führt zu einer weiteren Steigerung der Qualität ankommender, auswertbarer Sprachbefehle.

**[0019]** Weiterhin ist erfindungsgemäß vorgeschlagen, daß die Ermittlung der Richtcharakteristik und das Nachführen des virtuellen Mikrofonortes noch während eines einzelnen Sprechvorganges erfolgt. Eine solche Realisierung ist mit geeigneten elektronischen Mitteln möglich, die eine entsprechend schnelle Auswertbarkeit liefern.

**[0020]** Eine entsprechende Mehrfachmikrofonanordnung in erfindungsgemäßer Weise ist in den kennzeichnenden Merkmalen des Patentanspruches 5 angegeben. Auch hierbei steht im Vordergrund eine Mehrfachanordnung einzelner separater und autarker Mikrofone innerhalb des Fahrzeuges. Eine entsprechende Einzelsignalebewertung der Mikrofone erfolgt dabei in einer digitalen Signalprozessoreinheit, in welcher zunächst die ankommenden einzelnen Mikrofonsignale jeweils separat analysiert bzw. fouriertransformiert werden. Weiterhin werden analoge Größen der Einzelmikrofonsignale innerhalb dieser DSP (Digitaler Signal Prozessor) digitalisiert. Der so für jedes Einzelmikrofon vorliegende Parametersatz, der

jedoch auf ein und dieselbe Sprachquelle zurückgeht, können nun Laufzeit-, und Kohärenzbestimmung, zusätzlich zu einer Ortung des virtuellen Mikrofonortes, also des Ortes der Sprachquelle vorgenommen werden. Damit kann eine entsprechende virtuelle Richtung oder Nachführung bzw. eine daraufhin gerichtete Richtcharakteristik der gesamten Mehrfachmikrofonanordnung realisiert werden. Wichtig ist hierbei die simultane Bearbeitung aller einzelnen Signale, um eine gewisse Verarbeitungsschnelligkeit ermöglichen zu können. Hernach erfolgt eine digitale Nachbearbeitung und die entsprechende selektive Erkennung von darin enthaltenen Sprachbefehlen, die die Eingabeeinheit kennt. Darauf aufbauend werden die erkannten Sprachbefehle dann in einer entsprechenden Einheit generiert und somit elektrische Stellsignale, die aus den erkannten Sprachbefehlen ermittelt werden, den entsprechenden Stellorganen zugeführt.

**[0021]** In Anwendungen kann daran gedacht werden, Audio- Informations- und Kommunikationssysteme auf diese Weise zu steuern, aber auch fahrzeuginterne Einrichtungen, wie Navigationssystem, Klimaanlage, Beheizungssystem, Lichtanlage sowie ggf. auch Sicherheitseinrichtungen und auch Betriebszustände des Fahrzeuges. Bei den letztgenannten Einwirkungsmöglichkeiten handelt es sich um Einwirkungen auf das Betriebsverhalten des Fahrzeuges selbst. Hierbei ergibt sich selbstredend eine hohe Anforderung an die sichere Auswertung der Sprachbefehle, da sie ansonsten zu Fehlfunktionen führen können, die die Sicherheit des Fahrzeuges im Betrieb dramatisch beeinträchtigen können.

**[0022]** Die Erfindung ist in der Zeichnung dargestellt und nachfolgend näher beschrieben.

**[0023]** Die Abbildung zeigt eine schematische Darstellung im Hinblick auf die Funktionsweise des erfindungsgemäßen Verfahrens sowie der Mehrfachmikrofonanordnung selbst. Innerhalb des Fahrzeuges sind mehrere Mikrofone 1, 2, 3, ...räumlich verteilt an verschiedenen Orten angeordnet. Die räumlich verteilte Anordnung erfolgt dabei in geeigneter Weise so, daß sich keine planaren oder linearen Mikrofonarrays ausbilden. Vielmehr ist eine 3 dimensionale Verteilung der Mikrofone notwendig. Dies bedingt dann eine entsprechende Verteilung der einzelnen Mikrofone innerhalb der Fahrgastzelle insgesamt. Jedes der Mikrofone ist dabei signaltechnisch mit der digitalen Signalprozessoreinheit DSP 10 verbunden. Innerhalb der besagten digitalen Signalprozessoreinheit findet eine Sprachsignalanalyse statt zu welcher auch eine Fouriertransformation der einkommenden Einzelmikrofonsignale jeweils separat voneinander vorgenommen werden kann. Die Bildung eines Kohärenzbezuges aus der simultanen Bearbeitung aller Einzelmikrofonsignale gestattet zum ersten eine Signallaufmessung und die Inbezugsetzung aller Signale zueinander, so daß eine Ortung des virtuellen Mikrofonortes, d. h. der Ort der Sprachquelle möglich ist.

**[0024]** So entsteht zu jedem Einzelmikrofonsignal ein jeweils unabhängiger Satz von Parametern. Diese Parameter werden in der digitalen Signalprozessoreinheit DSP dahingehend ausgewertet, daß eine Ortung des virtuellen Mikrofonortes, also des Ortes der Sprachquelle und eine darauf abgestellte Richtcharakteristik virtuell ermittelbar ist. Diese Richtcharakteristik ist dabei darauf abgezielt, das Spracheingabesignal für die sprachgesteuerte Befehlseingabe in optimaler Weise erkennen zu können. Mit Hilfe der so ermittelten Richtcharakteristik ist selbst bei Variation des Ortes der Sprachquelle ein optimales Spracheingabesignal gewinnbar. Das heißt, der virtuelle Mikrofonort kann somit simultan, d. h. noch während des Sprechvorgangs so optimiert nachgeführt werden, daß eine optimale Erkennung der sprachgenerierten Befehlseingabe möglich ist.

**[0025]** In einer digitalen Nachbearbeitung 20 der Spracherkennung erfolgt somit eine Prüfung auf Plausibilität. Hier werden die aus den Einzelmikrofonsignalen und der daraus bestimmten Richtcharakteristik gewonnenen Sprachinformationen noch einmal selektiv bewertet und auf Plausibilität geprüft.

**[0026]** Damit ist realisiert, daß die simultane Verarbeitung aller Einzelmikrofonsignale mit diesem digitalen Signalprozessorsystem zu einem resultierenden Gesamtsignal führt. Dieses resultierende Gesamtsignal ist es dann, welches in der Spracherkennung ankommt und die Plausibilitätsprüfung auf die Verfügbarkeit entsprechender Befehle hin auch durchgeführt werden kann.

**[0027]** Das gesamte System ist somit und so zu sagen adaptiv, womit gemeint ist, daß die Bewegungen der Sprachquelle, d. h. beispielsweise des Kopfes des Sprechers, zeitsynchron detektiert werden. Mit dieser Kenntnis wird der Fokus der Mikrofongruppe adaptiv auf den Ort der Schallentstehung ausgerichtet und nachgeführt. Zu diesem einen, resultierenden Gesamtsignal hinzu kommend, erfolgt dann die Spracherkennung und die entsprechende Auswahl eines der verfügbaren Sprachbefehle. Dieser Sprachbefehl wird sodann einer Spracheingabeeinheit 30 zugeführt, in welchem die besagten dazugehörigen Stellsignale generiert werden. Mit Hilfe dieser generierten Stellsignale werden dann aus den erkannten Sprachbefehlen die entsprechenden Geräte und Systeme in entsprechender Weise angesteuert.

**[0028]** Insgesamt ergibt sich somit, daß zwei untergeordnete Verfahrensweisen hierbei das gesamte erfindungsgemäße Verfahren ergeben. Zum einen wird zunächst eine Mikrofonsignalauswertung aller Einzelmikrofonsignale durchgeführt, woraus zeitsynchron eine Lokalisierung des virtuellen Mikrofonortes ermittelt wird. Mit Hilfe dieser Ermittlung werden nun die ankommenden Signale dahingehend optimiert, daß sich ein resultierendes Gesamtsignal in optimierter Weise ergibt. Vor Ansteuerung der Befehlseingabe erfolgt in einem weiteren Verfahrensschritt eine Prüfung auf Plau-

sibilität.

**[0029]** Insgesamt ist das System so sicher ausgelegt, daß auch betriebsbezogene Einstellungen über sprachgenerierte Befehlseingaben zukünftig möglich sind, da das Verfahren ein optimales Maß an der Erkennung tatsächlich "gemeinter" Sprachbefehle durchstellen kann. Dies betrifft selbst solche, die das Fahrverhalten im Betrieb des Fahrzeuges beeinflussen, und somit an die Zuverlässigkeit und Genauigkeit der gesamten Spracherkennung hohe Anforderungen stellt. Darüber hinaus kann das System auch adaptiv angelegt sein, in dem Sinne, daß gewisse "stimmbezogenen" Parameter, die personenindividuell sind erkannt werden können, bspw. durch Mustervergleich. Dadurch kann das Spracheingabesystem selbsttätig Personen mit ihren Sprachspezifika erkennen, womit sich das System selbsttätig auf verschiedene Personen, bzw. verschiedene Fahrzeugführer einstellen. Hierzu müssen zu bereits beschriebenen elektronischen Komponenten Speicherelemente hinzukommen, mit deren Hilfe Sprachmuster bzw. Sprachspezifika abgespeichert, bzw. adaptiv abgespeichert werden können.

#### Patentansprüche

1. Verfahren zum Betrieb einer Mehrfachmikrofonanordnung in einem Kraftfahrzeug, bei welchem die Einzelsignale der Einzelmikrofone erfaßt und elektronisch nachbearbeitet werden, **dadurch gekennzeichnet**, daß zur präzisen Ansteuerung einer sprachgesteuerten Befehlseingabe von Steuer- oder Betätigungsbefehlen, oder zum präzisen Betrieb der Telefonanlage die Signale der Einzelmikrofone simultan und in Zeitbezug zueinander erfaßt und analysiert werden, und daß aus den daraus gewonnenen Parametern und dem Kohärenzbezug der Einzelsignale die Sprachsignale vor Ansteuerung der Befehlseingabe auf Plausibilität geprüft werden.
2. Verfahren nach Anspruch 1, **dadurch gekennzeichnet**, daß zur Sprachsignalanalyse die Signale der Einzelmikrofone zunächst durch A/D-Wandlung digitalisiert und dann fouriertransformiert werden.
3. Verfahren nach Anspruch 1 oder 2, **dadurch gekennzeichnet**, daß zusätzlich aus den Signalparametern und dem Kohärenzbezug eine Ortung durchgeführt wird, mit deren Hilfe die Richtung der auszuformenden Richtkeule bestimmt wird.
4. Verfahren nach Anspruch 3, **dadurch gekennzeichnet**, daß aus Abprüfen der Plausibilität und virtueller Lokalisierung der Sprachquelle, übrige delokale Geräuschquellen durch destruktive Interferenz minimiert werden.
5. Verfahren nach einem der vorhergehenden Ansprü-

che, **dadurch gekennzeichnet**, daß die Ermittlung der Richtcharakteristik und das Nachführen des virtuellen Mikrofonortes noch während eines einzelnen Sprechvorganges erfolgt.

5

6. Verfahren nach einem der vorhergehenden Ansprüche, **dadurch gekennzeichnet**, daß die Sprachsignale auch hinsichtlich erfaßbarer Stimmspezifika analysiert und die daraus ermittelbaren Parameter adaptiv abgespeichert und daraus eine individuelle Stimmerkennung von Personen durch Mustervergleich möglich ist. 10
  
7. Mehrfachmikrofonanordnung in einem Kraftfahrzeug, mit einer Mehrzahl innerhalb der Fahrgastzelle verteilt angeordneten Mikrofone sowie einer nachgeschalteten elektronischen Verarbeitungseinheit, **dadurch gekennzeichnet**, daß die Mikrofone (1, 2, 3, ...) signaltechnisch mit einer digitalen Signalprozessoreinheit DSP (10) derart verschaltet bzw. verbunden sind, daß jedes Einzelmikrofonsignal für sich simultan und zeitsynchron auswertbar ist, woraus eine Richtcharakteristik eines virtuellen Mikrofonortes ermittel- und nachführbar ist und daß in der Bewertung der Mikrofonensignale zu einem resultierenden Gesamtsignal die Sprachbefehle generiert und vor Auslösung innerhalb einer digitalen Nachbearbeitungseinheit (20) zur Spracherkennung auf Plausibilität prüfbar sind. 15  
20  
25  
30
  
8. Mehrfachmikrofonanordnung nach Anspruch 7, **dadurch gekennzeichnet**, daß die Einzelmikrofone (1, 2, 3, ...) räumlich d.h. unter Vermeidung linearer oder planarer Mikrofonarrays in der Fahrgastzelle des Kraftfahrzeuges verteilt angeordnet sind. 35
  
9. Mehrfachmikrofonanordnung nach einem der vorhergehenden Ansprüche, **dadurch gekennzeichnet**, daß die digitale Signalprozessoreinheit DSP (10) sowie die digitale Nachbearbeitung und Spracherkennung (20) und die Spracheingabeeinheit (30) zu einer zusammenhängenden Baugruppe bzw. zu einem zusammenhängenden Gerät zusammengefaßt sind. 40  
45
  
10. Mehrfachmikrofonanordnung nach einem oder mehreren der vorhergehenden Ansprüche, **dadurch gekennzeichnet**, daß digitale Speicher- mittel zur adaptiven Stimmerkennung und Stimm- charakteristikaerkennung vorgesehen sind. 50

55

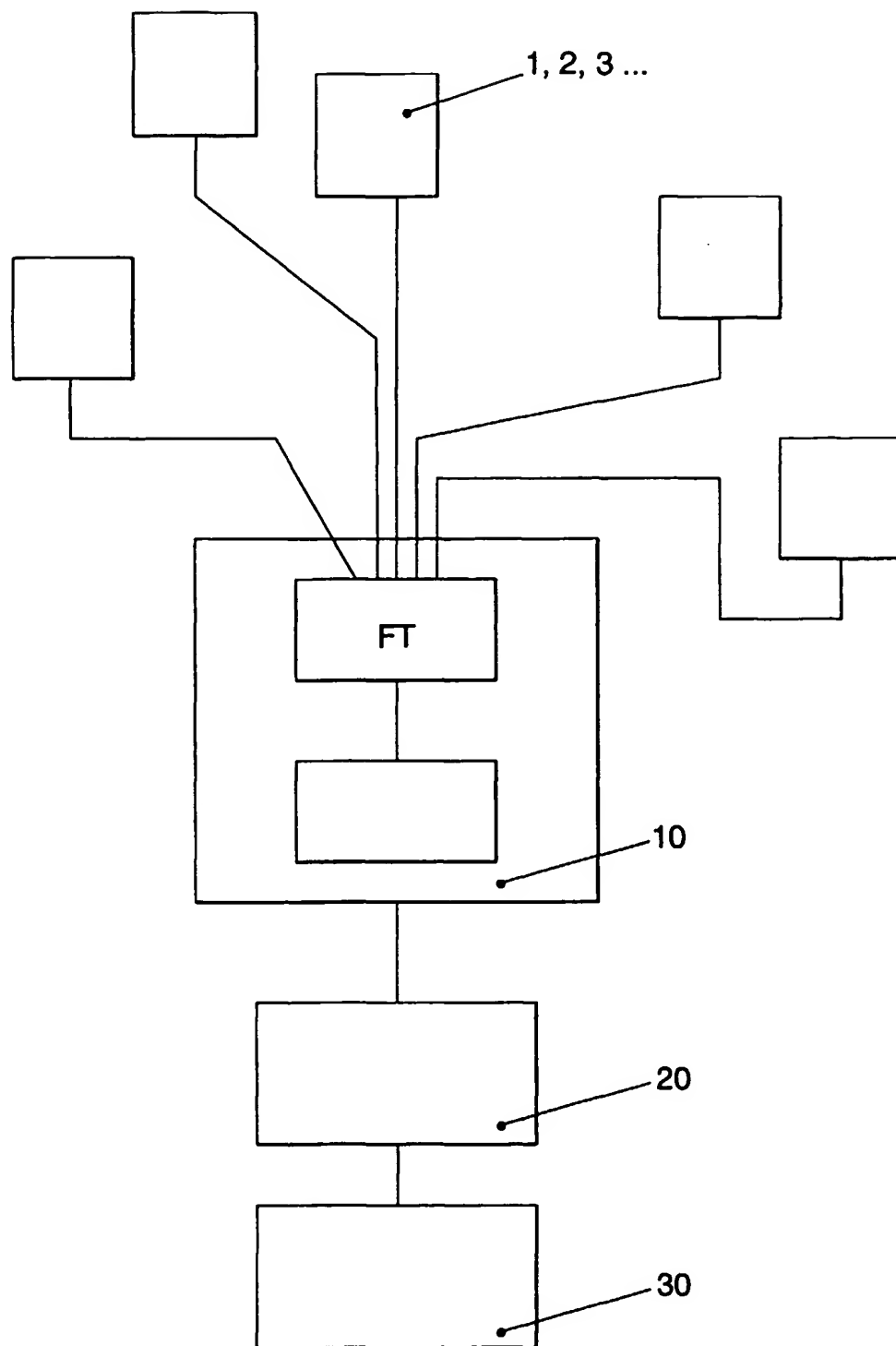


FIG. 1